UNITED STATES PATENT APPLICATION

for

A BULK NON-PLANAR TRANSISTOR HAVING STRAINED ENHANCED MOBILITY AND METHODS OF FABRICATION

Inventors:

Nick Lindert

Stephen M. Cea

Prepared by:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN
12400 Wilshire Boulevard
Seventh Floor
Los Angeles, CA 90025-1026
(408) 720-8300

Attorney Docket No.: __42P18257_____

A BULK NON-PLANAR TRANSISTOR HAVING STRAINED ENHANCED MOBILITY AND METHODS OF FABRICATION

BACKGROUND OF THE INVENTION

1. FIELD OF THE INVENTION

[0001] The present invention relates to the field of integrated circuit manufacturing and more particularly to the formation of a strain enhanced mobility bulk nonplanar transistor and its method of fabrication.

2. DISCUSSION OF RELATED ART

[0002] Modern integrated circuits, such as microprocessors, are made up of literally hundreds of millions of transistors coupled together. In order to improve the performance and power of integrated circuits, new transistor structures have been proposed. A nonplanar transistor, such as a tri-gate transistor, has been proposed to improve device performance. A tri-gate transistor 100 is illustrated in Figure 1A and 1B. Figure 1A is an illustration of a overhead/side view of a tri-gate transistor 100 and Figure 1B is an illustration of a cross-sectional view taken through the gate electrode of a tri-gate transistor 100. Tri-gate transistor 100 includes a silicon body 102 having a pair of laterally opposite sidewalls 103 and a top surface 104. Silicon body 102 is formed on an insulating substrate including an oxide layer 106 which in turn is formed on a monocrystalline silicon substrate 108. A gate dielectric 110 is formed on the top surface 104 and on the sidewalls 103 of silicon body 102. A gate electrode 120 is formed on the gate dielectric layer 110 and surrounds the silicon body 102. A pair of source/drain regions 130 are formed in the silicon body 102 along laterally opposite sidewalls of gate electrode 120. Transistor 130 can be said to be a tri-gate transistor because it essentially has three gates (G_1, G_2, G_3) which essentially form three transistors. Tri-gate transistor 100 has a first gate/transistor on one side 103 of silicon body 102, a second gate/transistor on a top surface 104 of silicon body

1

102 and a third gate/transistor on the second side 103 of silicon body 102. Each transistor provides current flow proportional to the sides of silicon body 102. The tri-gate transistor are attractive because they have large current per area which improves device performance.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] Figure 1A shows an overhead view of a standard tri-gate transistor.

[0004] Figure 1B shows a cross-sectional view of standard tri-gate transistor.

[0005] Figure 2 is an illustration of a bulk tri-gate transistor having a strain induced mobility in accordance with an embodiment with the present invention.

[0006] Figures 3A-3I illustrate a method of forming a bulk tri-gate transistor having a strain enhanced mobility in accordance with an embodiment of the present invention.

[0007] Figures 4A-4C illustrate a method of forming a bulk tri-gate transistor having a strain enhanced mobility in accordance with an embodiment of the present invention.

[0008] Figure 5 illustrates crystal lattices for a bulk silicon, a strained silicon germanium semiconductor body and a stained silicon capping layer.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0009] Embodiments of the present invention are bulk nonplanar transistors having strained enhanced mobility and their methods of fabrication. In the following description, numerous specific details have been set forth in order to provide a thorough understanding of the present invention. In other instances, well known semiconductor processing and fabrication techniques have not been set forth in particular detail in order to not unnecessarily obscure the present invention.

[0010] Embodiments of the present invention are bulk nonplanar transistors having strained enhanced mobility and their methods of fabrication. Embodiments of the present invention include a semiconductor body which places a capping layer formed on or around the semiconductor body under strain. A capping layer under strain increases the mobility of carriers in the device which increases the current of the device which can be used to improve circuit speeds.

[0011] An example of a bulk nonplanar or tri-gate transistor 200 having strain enhanced mobility is illustrated in Figure 2. Transistor 200 is formed on a bulk semiconductor substrate 202. In an embodiment of the present invention, the substrate 202 is a monocrystalline silicon substrate. Formed in semiconductor substrate 202 are a pair of spaced apart isolation regions 204, such as shallow trench isolation (STI) regions, which define the substrate active region 206 therebetween. Substrate 202, however, need not necessarily be a silicon monocrystalline substrate and can be other types of substrates, such as but not limited to germanium (Ge), silicon germanium (Si_xGe_y), gallium arsenide (GaAs), InSb, GaP, and GaSb. The active region 206 is typically doped to a p type conductivity level between 1×10^{16} to 1×10^{19} atoms/cm^3 for an n type device and doped to an n type conductivity level between 1×10^{16} to 1×10^{19} atoms/cm^3 for a p type device. In other embodiments of the present invention, the active region 206 can be an undoped semiconductor, such as an intrinsic or undoped silicon monocrystalline substrate.

[0012] Transistor 200 has a semiconductor body 208 formed on active substrate region 206 of bulk substrate 202. The semiconductor body 208 has a top surface 209 and a pair of laterally opposite sidewalls 211. The top surface 209 is separated from the bottom surface

4

formed on semiconductor substrate 206 by a distance which defines the body height. The laterally opposite sidewalls 211 of the semiconductor body 208 are separated by a distance which defines the body width. The semiconductor body 208 is a monocrystalline or single crystalline semiconductor film. In an embodiment of the present invention, the semiconductor body 208 is formed from a semiconductor material different than the semiconductor used to form the bulk substrate 202. In an embodiment of the present invention, the semiconductor body 208 is formed from a single crystalline semiconductor having a different lattice constant or size than the bulk semiconductor substrate 202 so that the semiconductor body 208 is placed under strain. In an embodiment of the present invention, the bulk semiconductor substrate is a monocrystalline silicon substrate and the semiconductor body 208 is a single crystalline silicon-germanium alloy. In an embodiment of the present invention, the silicon germanium alloy comprises between 5-40% germanium and ideally approximately between 15-25% germanium.

[0013] In an embodiment of the present invention, the bulk semiconductor substrate 202 is a monocrystalline silicon substrate and the semiconductor body 208 is a silicon-carbon alloy.

[0014] In an embodiment of the present invention, semiconductor body 208 is formed to a thickness less than the amount at which the exterior surfaces of the semiconductor body 208 will cause relaxation in the crystal lattice. In an embodiment of the present invention, semiconductor body 208 is formed to a thickness between 100-2000Å and more particularly between 200-1000Å. In an embodiment of the present invention, the thickness and height of the semiconductor body 208 are approximately the same.

[0015] In an embodiment of the present invention, the width of the semiconductor body 208 is between half the body 208 height to two times the body 208 height. In an embodiment of the present invention, semiconductor body 208 is doped to a p type conductivity with a concentration between 1×10^{16} to 1×10^{19} atoms/cm^3 for an n type semiconductor device and is doped to an n type conductivity with a concentration between 1×10^{16} to 1×10^{19} atoms/cm^3 for a p type semiconductor device. In an embodiment of the present invention, the semiconductor body 208 is intrinsic semiconductor, such as an undoped or intrinsic silicon film.

5

[0016] Transistor 200 includes a semiconductor capping layer 210 formed on the sidewalls 211 of semiconductor body 208 as well as on the top surface 209 of semiconductor body 208. Semiconductor capping layer 210 is a single crystalline semiconductor film. In an embodiment of the present invention, the semiconductor capping layer 210 is formed of a semiconductor material having a different lattice constant than the semiconductor body 208 so that a strain is formed in the capping layer. In an embodiment of the present invention, the capping layer has a tensile strain. A tensile strain is thought to improve the mobility of electrons. In an embodiment of the present invention, the capping layer has a compressive strain. A compressive strain is thought to improve hole mobility. In an embodiment of the present invention, current flows in a direction perpendicular to the strain in capping layer 210. In an embodiment of the present invention, the strain in the capping layer 210 on the sidewalls 211 of semiconductor body 208 is greater than the strain in the capping layer 210 on the top surface 209 of semiconductor body 208.

[0017] In an embodiment of the present invention, the semiconductor capping layer 210 is a single crystalline silicon film. In an embodiment of the present invention, the capping layer 210 is a single crystalline silicon film formed on a silicon-germanium alloy body 208. A single crystalline silicon film formed on a silicon-germanium alloy semiconductor body 208 will cause the single crystalline silicon film to have a tensile stress. In an embodiment of the present invention, the capping layer 210 is a single crystalline silicon film formed on a silicon-carbon alloy semiconductor body 208. A single crystalline silicon capping layer 210 formed on a silicon-carbon alloy semiconductor body 208 will cause the single crystalline silicon film 210 to have a compressive stress.

[0018] In an embodiment of the present invention, the semiconductor capping layer 210 is formed to a thickness less than the amount at which the lattice of the single crystalline film will relax. In an embodiment of the present, the semiconductor capping layer 210 is formed to a thickness between 50-300Å. In an embodiment of the present invention, the thickness of the capping layer on the sidewalls 211 of semiconductor body 208 is the same as the thickness of the capping layer 210 on the top surface 209 of semiconductor body 208 as illustrated in Figure 2. In an embodiment of the present invention, the semiconductor

6

capping layer 210 is formed thicker on the top surface of the semiconductor body 208 than on the sidewalls 211, such as shown, for example, in Figure 4C.

[0019] Transistor 200 includes a gate dielectric layer 212. Gate dielectric layer 212 is formed on capping layer 210 formed on the sidewalls 211 of semiconductor body 208 and is formed on semiconductor capping layer 210 formed on the top surface 209 of semiconductor body 208. Gate dielectric layer 210 can be any well known gate dielectric layer. In an embodiment of the present invention, the gate dielectric layer is a silicon dioxide (SiO_2), silicon oxynitride (SiO_xN_y), or a silicon nitride (Si_3N_4) dielectric layer. In an embodiment of the present invention, the gate dielectric layer 212 is a silicon oxynitride film formed to a thickness between 5-20Å. In an embodiment of the present invention, the gate dielectric layer 212 is a high K gate dielectric layer, such as a metal oxide dielectric, such as but not limited to tantalum pentaoxide (Ta_2O_5), titanium oxide (TiO_2), hafnium oxide (HfO) and zirconium oxide (ZrO). Gate dielectric layer 212, however, can be other types of high K dielectrics, such as but not limited to PZT and BST.

[0020] Transistor 200 includes a gate electrode 214. Gate electrode 214 is formed on and around the gate dielectric layer 212 as shown in Figure 2. Gate electrode 214 is formed on and adjacent to gate dielectric layer 212 formed on capping layer 210 formed on sidewall 211 of semiconductor body 208 and is formed on gate dielectric layer 212 formed on capping layer 210 formed on the top surface 209 of semiconductor body 208 and is formed on or adjacent to gate dielectric layer 212 formed on capping layer 210 formed on sidewall 211 of gate electrode 208 as shown in Figure 2. Gate electrode 214 has a pair of laterally opposite sidewalls 216 separated by a distance which defines the gate length (Lg) of transistor 200. In an embodiment of the present invention, the laterally opposite sidewalls 216 of gate electrode 214 run in a direction perpendicular to the laterally opposite sidewalls 211 of semiconductor body 208. Gate electrode 214 can be formed of any suitable gate electrode material. In an embodiment of the present invention, gate electrode 214 comprises polycrystalline silicon film doped to a concentration density between 1×10^{19} to 1×10^{20} atoms/cm^3. Gate electrode 214 can be doped to an n type conductivity for an n type device and p type conductivity for a p type device. In an embodiment of the present invention, the gate electrode can be a metal gate electrode. In an embodiment of the

7

present invention, the gate electrode 214 is formed of a metal film having a work function which is tailored for an n type device, such as a work function between 3.9 eV to 4.2 eV. In an embodiment of the present invention, the gate electrode 214 is formed from a metal film having a work function tailored for a p type device, such as a work function between 4.9 eV to 5.2 eV. In an embodiment of the present invention, the gate electrode 214 is formed from a material having midgap work function between 4.6 to 4.8 eV. A midgap work function is ideal for use when semiconductor body 208 and capping layer 210 are intrinsic semiconductor films. It is to be appreciated that gate electrode 214 need not necessarily be a single material and can be composite stack of thin films, such as but not limited to polycrystalline silicon/metal electrode or metal polycrystalline silicon electrode.

[0021] Transistor 200 has a pair source/drain regions formed in semiconductor body 208 as well as in capping layer on opposite sides of a laterally opposite sidewalls 216 of gate electrode 214 as shown in Figure 2. The source/drain regions 218 are doped to an n type conductivity when forming an n type device and doped to a p type conductivity when forming a p type device. In an embodiment of the present invention, the source/drain regions have doping concentration of between 1×10^{19} to 1×10^{21} atoms/cm^3. The source/drain regions 218 can be formed of uniform concentration or can include subregions of different concentrations or doping profiles, such as tip regions (e.g., source/drain extensions). In an embodiment of the present invention, when transistor 200 is a symmetrical transistor the source and drain regions will have the same doping concentration profile. In an embodiment of the present invention, transistor 200 is an asymmetrical transistor, the source region and drain region may vary in order to obtain particular electrical characteristics.

[0022] The portion of the semiconductor body 208 and capping layer 210 located between the source/drain regions 216 and beneath the gate electrode 214 defines a channel region of the transistor. The channel region can also be defined as the area of semiconductor body 208 and capping layer 210 surrounded by gate electrode 214. The source/drain regions typically extend slightly beneath the gate electrode through, for example, diffusion to define the channel region slightly smaller than the gate electrode length (Lg). When transistor 300 is turned "ON" an inversion layer is formed in the channel region of the

device which forms a conductive channel which enables current to travel between the source/drain region 340. The inversion layer or conductive channel forms in the surface of the capping layer on the sidewalls 211 of semiconductor body 208 as well as in the surface of capping layer 210 on the top surface 209 of semiconductor body 208.

[0023] By providing a gate dielectric layer 212 and a gate electrode 214 which surrounds the semiconductor body 208 and capping layer 210 on three sides, the nonplanar transistor is characterized as having three channels and three gates, one gate (G1) which extends between the source/drain regions on one side 211of semiconductor body 208, a second gate (G2) which extends between the source/drain regions on the top surface 209 of semiconductor body 208 and the third (G3) which extends between the source/drain regions on sidewall 211 of semiconductor body 208. The gate "width" (Gw) of transistor 200 is the sum of the width of the three channel regions. That is, the gate width of transistor 200 is equal to the height of semiconductor body 208 plus the thickness of the capping layer on the top surface of sidewall 211, plus the width of semiconductor body 208 plus the thickness of the capping layer on each of the sides 211 of semiconductor body plus the height of semiconductor body 208 plus the thickness of capping layer 210 on the top surface 209 of semiconductor body 208. Larger "width" transistor can be obtained by using multiple semiconductor bodies 208 and capping layers surrounded by a single gate electrode, such as illustrated in Figure 3I.

[0024] Although a tri-gate transistor 200 is illustrated in Figure 2, the present invention is equally applicable to other nonplanar transistors. For example, the present invention is applicable to a "finfet" or a double gate transistor or just two gates are formed on opposite sides of the semiconductor body. Additionally, the present invention, is applicable to "omega" gates or wrap around gate devices where the gate electrode wraps around the semiconductor body as well as underneath a portion of the semiconductor body. Performance of "finfet" devices and "omega" devices can be improved by including a strained capping layer 210 formed on a semiconductor body 208 and thereby enhancing the mobility of carriers in the device. It is to be appreciated that a nonplanar device is a device which when turned "ON" forms a conductive channel or a portion of the conductive channel in a direction perpendicular to the plane of the substrate 202. A nonplanar

9

transistor can also be said to be a device where the conductive channel regions are formed both in the horizontal and vertical directions.

[0025] Figures 3A-3I illustrate a method of forming a bulk nonplanar transistor having a strain enhanced mobility in accordance with an embodiment of the present invention. First a semiconductor substrate 300 is provided as shown in Figure 3A. In an embodiment of the present invention, semiconductor substrate 300 is a monocrystalline silicon substrate. Substrate 300 need not necessarily be a silicon substrate and can be other types of substrates, such as a silicon germanium substrate, a germanium substrate, a silicon germanium alloy, a gallium arsenide, InSb, and GaP. In an embodiment of the present invention, the semiconductor substrate 300 is an intrinsic (i.e., undoped) silicon substrate. In other embodiments of the present invention, the semiconductor substrate 300 is doped to a p type or n type conductivity with a concentration between 1×10^{16} to 1×10^{19} atom/cm^3. Next, a mask having mask portions 302 for forming isolation regions is formed on substrate 300 as shown in Figure 3A. In an embodiment of the present invention, the mask is an oxidation resistant mask. In an embodiment of the present invention, the mask portions 302 comprise a thin pad oxide layer 304 and a thicker silicon nitride or oxidation resistant layer 306. The mask portions 302 define active regions 308 in substrate 300 where transistor bodies are to be formed. The mask portions 302 can be formed by blanket depositing a pad oxide layer and then a silicon nitride layer over substrate 300. Next, well known photolithography techniques are used to mask, expose and develop a photoresist masking layer over locations where mask portions 302 are to be formed. The nitride film 306 and the pad oxide layers 304 are then etched in alignment with the formed photoresist mask to form mask portions 302 as shown in Figure 3A.

[0026] In an embodiment of the present invention, mask portions 302 have a width (W1) which is the minimum width or minimum feature dimension (i.e., critical dimension (CD)) which can be defined utilizing photolithography in the fabrication of the transistor. Additionally, in an embodiment of the present invention, mask portions 302 are separated by a distance D1 which is the minimum distance which can be defined utilizing photolithography in the fabrication process. That is, mask portions 302 have the smallest dimension and are spaced apart by the smallest dimension (i.e., critical dimensions) which

can be reliably and achieved utilizing the photolithography process used to fabricate the transistor. In this way, mask portions 302 are defined to have the smallest size and greatest density capable of being achieved with the photolithography process used in fabrication of the transistor.

[0027] In an embodiment of the present invention, mask portions 302 have a thickness (T1) which is equal to or greater than the thickness or height desired for the subsequently formed semiconductor body or bodies.

[0028] Next, as shown in Figure 3B, the exposed portions of semiconductor 300 are etched in alignment with the outside edges of mask portion 302 to form trench openings 310. The trench openings are etched to a depth sufficient to isolate adjacent transistors from one another.

[0029] Next, as shown in Figure 3C, the trenches are filled with a dielectric layer 312 to form shallow trench isolation (STI) regions 312 in substrate 300. In an embodiment of the present invention, the dielectric layer is formed by first growing a thin liner oxide in the bottom of sidewalls of trench 310. Next, trench 312 is filled by blanket depositing an oxide dielectric layer over the liner oxide by, for example, a high density plasma (HDP) chemical vapor deposition process. The fill dielectric layer will also form on the top of mask portions 302. The fill dielectric layer can then be removed from the top of mask portions 302 by, for example, chemical mechanical polishing. The chemical mechanical polishing process is continued until the top surface of mask portions 302 is revealed and the top surface of shallow trench isolation regions 312 substantially planar with the top surface of mask portion 302 as shown in Figure 3C.

[0030] Although shallow trench isolation regions are ideally used in the present invention, other well known isolation regions and techniques, such as local oxidation of silicon (LOCOS) or recessed LOCOS may be utilized.

[0031] Next, as shown in Figure 3D, mask portions 302 are removed from substrate 300 to form semiconductor body openings 314. First a silicon nitride portion 306 is removed utilizing an etchant which etches away the oxidation resistant or silicon nitride portion 306 without significantly etching the isolation regions 312. After removing silicon nitride portion 306, the pad oxide portion 304 is removed. Pad oxide portion 304 can be removed,

11

for example, with a wet etchant comprising hydrofluoric acid (HF). Removing of mask portions 302 forms a semiconductor body opening or trench 314 having substantially vertical sidewalls. The vertical sidewall enables the semiconductor body to be grown within the trench and confined therein to enable a semiconductor body to be formed with nearly vertical sidewalls.

[0032] Next, as shown in Figure 3E, a semiconductor body film 316 is formed in opening 314 as shown in Figure 3E. In an embodiment of the present invention, the semiconductor body film 316 is an epitaxial semiconductor film. In an embodiment of the present invention, when a strain enhanced semiconductor device is desired, the semiconductor film is formed from a single crystalline semiconductor film having a different lattice constant or different lattice size than the underlying semiconductor substrate upon which it is grown, so that the semiconductor film is under strain. In an embodiment of the present invention, the single crystalline silicon film 316 has a larger lattice constant or lattice size than the underlying semiconductor substrate 300. In an embodiment of the present invention, the single crystalline semiconductor film 316 has a smaller lattice size or constant than the underlying semiconductor substrate 300.

[0033] In an embodiment of the present invention, the semiconductor film 316 is an epitaxial silicon germanium alloy film selectively grown on a silicon monocrystalline substrate 300. A silicon germanium alloy can be selectively grown in an epitaxial reactor utilizing a deposition gas comprising, dichlorosilane (DCS), H_2, germane (GeH_4), and HCl. In an embodiment of the present invention, the silicon germanium alloy comprises between 5-40% germanium and ideally between 15-25% germanium. In an embodiment of the present invention, epitaxial semiconductor film 316 is a single crystalline silicon carbon alloy formed on a silicon substrate 300. The single crystalline semiconductor film 316 is deposited to a thickness desired for the thickness of the semiconductor body. In an embodiment of the present invention, it is grown or deposited to a thickness less than the height of the top surface of isolation regions 312. In this way, the isolation regions 312 confines the semiconductor film 316 within the trench so that a semiconductor film with nearly vertical sidewalls is formed. Alternatively, semiconductor film 316 can be blanket deposited over substrate 300 including within trench 314 and on top of isolation regions

12

312 and then polished back so that the semiconductor film 316 is removed from the top of the isolation regions and remains only within trenches 314 as shown in Figure 3E.

[0034] In an embodiment of the present invention, the semiconductor film 316 is an undoped or intrinsic semiconductor film. In an embodiment of the present invention, when fabricating a p type device, the semiconductor film 316 doped to an n type conductivity with a concentration between 1×10^{16} to 1×10^{19} atoms/cm^3. In an embodiment of the present invention, when fabricating an n type device the semiconductor film 316 is doped to a p type conductivity with a concentration between 1×10^{16} to 1×10^{19} atoms/cm^3. The semiconductor film 316 can be doped during deposition in an "insitu" process by including a dopant gas in the deposition process gas mix. Alternatively, the semiconductor film 316 can be subsequently doped by, for example, ion implantation or thermal diffusion to form a doped semiconductor film 316.

[0035] Next, isolation regions 312 are etched back or recessed to expose the sidewalls 320 of semiconductor film 316 and thereby form semiconductor bodies 318 as shown in Figure 3F. Semiconductor bodies 318 have nearly vertical sidewalls 320 because semiconductor film 316 was laterally confined by isolation regions 312 during deposition. Isolation regions 312 are etched back with an etchant which does not significantly etch the semiconductor film 316. When semiconductor film 316 is a silicon or silicon alloy isolation regions 312 can be recessed utilizing a wet etchant comprising HF. In an embodiment of the present invention, isolation regions are etched back to a level so that they are substantially planar with the top surface of the active regions 308 formed in semiconductor substrate 300 as shown in Figure 3F.

[0036] Next, as shown in Figure 3G, a semiconductor capping layer 322 is formed on the top surface 319 and sidewalls 320 of semiconductor body 318. Semiconductor capping layer 322 is a single crystalline semiconductor film. In an embodiment of the present invention, the semiconductor capping layer 322 is formed of a material having a different lattice constant or size than semiconductor body 318. In an embodiment of the present invention, semiconductor capping layer 322 is a single crystalline silicon film. In an embodiment of the present invention, semiconductor capping layer 322 is a single crystalline silicon film formed on a silicon germanium alloy body 318. In an embodiment

13

of the present invention, semiconductor capping layer 322 is a single crystalline silicon film formed on a silicon-carbon alloy semiconductor body 318. A single crystalline silicon capping layer 322 can be selectively deposited in an epitaxial deposition reactor utilizing a process gas comprising DCS, HCl and H_2. In an embodiment of the present invention, semiconductor capping layer 322 is formed to a thickness less than an amount which will cause substantial relaxation in semiconductor capping layer 322. In an embodiment of the present invention, semiconductor capping layer 322 is formed to a thickness sufficient to enable the entire inversion layer to be formed in the capping layer when the transistor is turn "ON". In an embodiment of the present invention, semiconductor capping layer 322 is formed to a thickness between 50-300Å. In an embodiment of the present invention, semiconductor capping layer 322 is an undoped or intrinsic semiconductor film. In an embodiment of the present invention, semiconductor capping layer 322 is doped to an n type conductivity between 1×10^{16} to 1×10^{19} atoms/cm^3 when forming a p type device and is doped to a p type conductivity between 1×10^{16} to 1×10^{19} atoms/cm^3 when forming an n type device. In an embodiment of the present invention, semiconductor capping layer 322 is doped in an insitu deposition process. Alternatively, capping layer 322 can be doped by other well known techniques, such as by ion implantation or solid source diffusion.

[0037] Next, as shown in Figure 3H, a gate dielectric film 324 is formed on capping layer 322 formed on the sidewalls 320 of semiconductor body 318 and is formed on the capping layer 322 formed on the top surface 319 of semiconductor body 318 as shown in Figure 3H. In an embodiment of the present invention, gate dielectric layer 324 is a grown gate dielectric layer, such as but not limited to a silicon dioxide layer, a silicon oxynitride layer or a combination thereof. A silicon oxide or silicon oxynitride layer can be grown on semiconductor capping layer utilizing a well known dry/wet oxidation process. When gate dielectric layer 324 is grown it will form only on semiconductor containing areas, such as capping layer 322 and not on isolation regions 312. Alternatively, gate dielectric layer 324 can be a deposited dielectric layer. In an embodiment of the present invention, gate dielectric layer 324 is a high K gate dielectric layer, such as a metal oxide dielectric layer, such as but not limited to hafnium oxide, zirconium oxide, tantalum oxide and titanium oxide. A high K metal oxide dielectric layer can be deposited by any well known

14

technique, such as chemical vapor deposition or sputter deposition. When gate dielectric layer 324 is deposited it will also form on isolation regions 312 .

[0038] Next, as shown in Figure 3H, a gate electrode material 326 is blanket deposited over substrate 300 so that it deposits onto and around gate dielectric layer 324. That is, the gate electrode material is deposited onto the gate dielectric layer 324 formed on capping layer 322 formed on the top surface of semiconductor body 318 and is formed or adjacent to capping layer 322 formed on the sidewalls 320 of semiconductor body 318. In an embodiment of the present invention, the gate electrode material 326 is polycrystalline silicon. In an embodiment of the present invention, the gate electrode material 326 is a metal film. In an embodiment of the present invention, gate electrode material 326 is a metal film having a work function tailored for an n type device and in an embodiment of the present invention, the gate electrode material is metal film having a work function tailored for a p type device. Gate electrode material 326 is formed to a thickness sufficient to completely cover or surround semiconductor bodies 318, capping layer 322 and gate dielectric layer 324 as shown in Figure 3H.

[0039] Next, as shown in Figure 3I, the gate electrode material 326 and gate dielectric layer 324 are patterned by well known techniques to form a gate electrode 330 and a gate dielectric layer 328. Gate electrode material 326 and gate dielectric layer 324 can be patterned utilizing well known photolithography and etching techniques. Gate electrode 330 has a pair of laterally opposite sidewalls 332 which define the gate length of the device. In an embodiment of the present invention, laterally opposite sidewalls 332 run in a direction perpendicular to semiconductor bodies 318. Although, a subtractive process is shown for the formation of gate electrode 330, other well known techniques, such as a replacement gate process may be utilized to form gate electrode 330.

[0040] Next, as also shown in Figure 3I, a pair of source/drain regions 340 are formed in capping layer 332 and semiconductor body 318 on opposite sides of gate electrode 330. When forming an n type device, source/drain regions can be formed to an n type conductivity with a concentration between 1×10^{20} to 1×10^{21} atoms/cm^3. In an embodiment of the present invention, when forming a p type device, source/drain regions having a p type conductivity with a concentration between 1×10^{20} to 1×10^{21} atoms/cm^3 can be formed.

15

Any well known technique, such as ion implantation or thermal diffusion, may be utilized to form the source/drain regions. When ion implantation is used, the gate electrode 330 can be used to mask the channel region of the transistor from the ion implantation process and thereby self-aligning the source/drain regions 340 with the gate electrode 330. Additionally, if desired, source/drain regions may include sub-regions, such as source/drain extensions and source/drain contact regions. Well known processes including formation of spacers can be utilized to form the sub-regions. Additionally, if desired, silicide can be formed on the source/drain regions 340 and on top of the gate electrode 330 to further decrease the electrical contact resistance. This completes the fabrication of bulk nonplanar transistor having strain enhanced mobility.

[0041] Well known "back end" techniques can be utilized to form metal contacts, metallization layers and interlayer dielectrics to interconnect various transistors together into functional integrated circuits, such as microprocessors.

[0042] A valuable aspect of the present invention, is that the capping layer increases the gate width of the transistor. In this way, minimum feature dimension and spacing can be used to form the semiconductor bodies and then the capping layer can be formed on and around the minimally defined semiconductor bodies to increase the gate width of the device. This increases the current per area of the device which improves device performance. Formation of a capping layer on minimally defined and separated features reduces the distance between minimally spaced bodies to a distance less than the critical dimension or less than the dimension achievable with photolithography process used to define the device. In this way, the formation of a capping layer enables larger gate width to be achieved with each semiconductor body while still defining the bodies with the minimum critical dimensions (CD) and spacing. Utilizing a capping layer to increase the gate width is valuable even in applications which do not require or desire stress enhanced mobility. As such, embodiments of the present invention include applications where, for example, silicon capping layers are formed on minimally spaced silicon bodies in order to increase the gate width of the fabricated transistor. Additionally, use of a capping layer to increase gate width per area is useful in non-bulk devices, such as tri-gate or nonplanar devices formed on insulated substrates, such as in silicon on insulator (SOI) substrates.

16

[0043] In embodiments of the present invention, stacks of semiconductor films (i.e., bulk semiconductor 300, semiconductor body 318 and capping layer 322) are engineered to produce high strain in the capping layer 322 which can dramatically increase carrier mobility. Figure 5 illustrates how a bulk silicon monocrystalline silicon substrate, a silicon germanium alloy semiconductor body 320 and a silicon capping layer 322 can produce high tensile stress in the silicon capping layer 322. When growing an epitaxial silicon germanium alloy film 316 on a monocrystalline substrate 300 (Figure 3E) the lattice constant of the plane 502 of the silicon germanium film 318 parallel to the surface of the silicon monocrystalline substrate 300 is matched to the silicon lattice of the bulk silicon substrate 300. The lattice constant of the plane 504 of the silicon germanium alloy 316 perpendicular to the silicon substrate surface is larger than the plane 502 parallel to the silicon substrate 300 due to the tetragonal distortion of the silicon germanium epitaxial film 316. Once the isolation regions 312 are recessed (Figure 3F) to form silicon germanium body 318 the silicon germanium lattice on the top 319 will expand and the lattice constant on the sides will contract due to the presence of free surface. In general the lattice constant on the sidewall 320 of the silicon germanium alloy 318 will be larger than the lattice constant on the top surface 319 of the silicon germanium alloy which will be greater than the lattice constant of the silicon germanium alloy on the silicon monocrystalline substrate. When a silicon capping layer 322 is grown on the strained silicon germanium alloy, (Figure 3G) the silicon germanium alloy 318 will impose its lengthened vertical cell dimension 504 on an already smaller cell dimension of the silicon capping layer 322 producing a orthorhombic strained silicon capping layer 322 on the sidewalls of the SiGe body 318. Thus, the silicon capping layer formed on the sidewalls 322 of the silicon germanium alloy will witness a substantial tensile strain and a lower but significant tensile strain on the top surface 319 of the silicon germanium alloy. The strain produced in silicon capping layer 322 is in a direction perpendicular to current flow in the device.

[0044] Figures 4A-4C illustrate a method of forming a bulk nonplanar transistor having strain enhanced mobility wherein the capping layer is formed thicker on the top surface of the semiconductor body than on the sidewalls. As illustrated in Figure 4A, semiconductor body film 316 is grown between isolation regions 312 as described with respect to Figure

17

3E. In this embodiment, however, a first portion 410 of the capping layer is grown on semiconductor body 316 prior to recessing isolation regions 312. In an embodiment of the present invention, silicon nitride layer 306 is formed thicker than necessary for the semiconductor body 318 so that additional room is provided to enable the first portion 410 of the semiconductor capping layer to be grown within the trench 310. In this way, the first portion of the capping layer 410 can be confined within the isolation regions 312. After formation of the first portion 410 of the capping layer, the isolation regions 312 are recessed back as described above to form a semiconductor body 318 having a capping layer 410 formed on the top surface thereof as shown in Figure 4B. Next, as shown in Figure 4C, a second portion 412 of the capping layer is grown on the sidewalls 320 of the semiconductor body 318 and on the first portion 410 of the capping layer formed on the top surface 319 of semiconductor body 320. In an embodiment of the present invention, the semiconductor capping layer 410 is formed to a thickness substantially equal to the thickness of the second portion of the capping layer 412. In this way, when a substantially square semiconductor body 318 is formed, the semiconductor body 318 plus capping layer will still provide a substantially square capped body. Next, processing can continue as illustrated in Figures 3H and 3I to complete fabrication of the bulk nonplanar transistor having a strain enhanced mobility.